Surpassing Simple Aggregation: Advanced Strategies for Analyzing Contextual-Level Outcomes in Multilevel Models

Becker, Dominik; Breustedt, Wiebke; Zuber, Christina Isabel

methods, data, analyses | Vol. 12(2), 2018, pp. 233-264 DOI: 10.12758/mda.2017.05

Surpassing Simple Aggregation: Advanced Strategies for Analyzing Contextual-Level Outcomes in Multilevel Models

Dominik Becker[1], *Wiebke Breustedt*[2] & *Christina Isabel Zuber*[3]

[1] *University of Tuebingen*
[2] *University of Duisburg-Essen/University of Cologne*
[3] *University of Konstanz*

Abstract

This article introduces two advanced analytical strategies for analyzing contextual-level outcomes in multilevel models: the multilevel SEM and the two-step approach. Since these strategies are seldom used in comparative survey research, we first discuss their methodological and statistical advantages over the more commonly applied approach of group mean aggregation. We then illustrate these advantages in an empirical analysis of the effect of citizens' support for democratic values at the individual level on a contextual-level outcome – the persistence of democracy – drawing on data from the World Values Survey and the Quality of Government project. Whereas we found no significant effect of support for democratic values in the model using simple group mean aggregation, citizens' support for democratic values was a significant predictor of democracies' estimated survival rate when applying latent aggregation in multilevel SEM and the two-step approach. The article corroborates previous concerns with simple aggregation and demonstrates how researchers can improve the validity of their analyses of contextual-level outcomes by using alternative strategies of aggregation.

Keywords: transformational mechanisms, contextual level outcomes, multilevel analysis, sampling error, democratic stability, democratic values

Despite significant methodological advancements, comparative social scientists continue to face the question of how to adequately test theoretical multilevel models empirically. Hierarchical modeling has evolved into a canonical statistical technique for regressing an individual-level variable on individual- and contextual-level predictors. There is no agreement when it comes to multilevel models where the dependent variable is analytically located on the contextual level, though.

Many comparative studies 'solve' this problem through measures of central tendency – such as the average – or the distribution of the data – such as percentages. They then use these aggregates as predictors for the contextual-level dependent variable (for examples, see Fails & Pierce, 2010; Lim, Bond, & Bond, 2005; Muller & Seligson, 1994). This approach has been criticized on both statistical and methodological grounds. Croon and van Veldhoven (2007) demonstrated that group mean aggregation may lead to biased estimates. Griffin (1997) argued that the aggregation procedure needs to take into account the complex theoretical relationships of independent variables at different levels of analysis. When applying simple aggregation, researchers may run the risk of drawing invalid conclusions about how individual-level predictors affect contextual-level outcomes (Snijders & Bosker, 1999).

Given these criticisms, researchers have proposed two more advanced strategies for analyzing contextual-level outcomes in multilevel models: the multilevel SEM and the two-step approach. Since multilevel SEM and the two-step approach are seldom used in comparative survey research, the article seeks to motivate researchers to improve the validity of their inferences when analyzing contextual-level outcomes by going beyond simple aggregation. In the following section, we introduce the methodological and statistical advantages of these two alternative techniques over the group means approach. In our analysis, we illustrate these advantages in an empirical study of the effect of citizens' support for democratic values at the individual level on a contextual outcome – the persistence

Acknowledgments

This paper benefited from feedback on several occasions including the SOCLIFE Graduate School at the University of Cologne, the annual conference of the European Survey Research Association (ESRA) in Lausanne and the joint conference of the International Political Science Association (IPSA) and the International Association for Political Science Students (IAPSS). In addition, we would like to thank Andreas Hartung and Steffen Hillmert for their helpful comments on our event-history models. Finally, we would like to thank two anonymous reviewers whose valuable suggestions helped to improve the overall quality and consistency of our manuscript.

Direct correspondence to

Dominik Becker, University of Tuebingen, Institute of Sociology, Wilhelmstr. 36, 72074 Tuebingen, Germany

E-mail: dominik.becker@ifsoz.uni-tuebingen.de

of democracy. We draw on data from the World Values Survey and the Quality of Government project and study 98 countries between 1946 and 2014. We compare the regression coefficients and confidence intervals of our individual-level predictor – support for democratic values – on democracies' persistence when applying the three methods. Whereas we found no significant effect of support for democratic values in the model using simple group mean aggregation, citizens' support for democratic values was a significant predictor of democracies' estimated survival rate when applying multilevel SEM and the two-step approach. In the final section we therefore conclude that comparative researchers who use simple group mean aggregation when regressing a contextual outcome on individual level predictors may run the risk of wrongly rejecting their hypothesis of interest.

Methodological Foundation and Statistical Background

Testing theoretical multilevel models with contextual-level outcomes poses two challenges. From a methodological point of view, researchers need to establish close correspondence between the theoretical multilevel mechanism and its empirical measurement. From a statistical perspective, they need to choose a method that is both valid and reliable for aggregating the individual-level predictors. In the following, we discuss the methodological foundations of multilevel analysis of macro-level social phenomena. We then proceed to introduce and compare three analytical strategies for analyzing contextual level outcomes: simple manifest group mean aggregation, latent aggregation through multilevel SEM, and the two-step approach. The results of the comparison are summarized in Table 1 at the end of this chapter.

Methodological Foundation

According to the paradigm of structural individualism (Udehn, 2002), the ultimate goal of the social sciences is to explain social phenomena on the contextual – or macro – level as a consequence of individuals' social actions on the individual – or micro – level. Structural individualism distinguishes three explanatory mechanisms (see Figure 1) (Hedström & Swedberg, 1998; Tranow, Beckers, & Becker, 2016). Situational mechanisms (1) link the objective characteristics of the social situation to the subjective expectations and evaluations of individuals. Action-formation mechanisms (2) explain individuals' actions given their subjective definition of the situation. This is a pure micro-level explanatory step. Transformational mechanisms (3) reconstruct how individuals' actions aggregate to create a new social situation. They thereby re-link the micro level to the macro level.

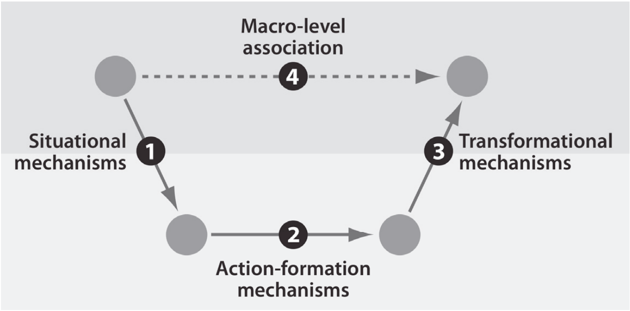

Figure 1 The social mechanisms of social science explanations.
Source: Hedström & Ylikoski (2010, p. 53)

Studying these theoretical mechanisms empirically is not straightforward. Multilevel modeling (Bryk & Raudenbush, 1992; Hox, 2010) is a well-established statistical tool for testing situational and action formation mechanisms, that is, explanations that link social situations to individuals' expectations, evaluations, and actual decisions (Becker, Beckers, Franzmann, & Hagenah, 2016). By contrast, micro-to-macro (or, more technically, level-one to level-two) explanations constitute a blind spot of conventional multilevel analysis (henceforth MLA)[1] as transformational mechanisms are more difficult to analyze empirically (Opp, 2011; Raub, Buskens, & van Assen, 2011).

Three Analytical Strategies

The simple group means approach

When studying multilevel models with contextual-level outcomes, a common approach (Lim et al., 2005) is to aggregate all level-one variables (hereafter L1) to level-two variables (hereafter L2) by computing their group-specific arithmetic means. This manifest aggregation is followed by an L2-only regression (see Figure 2).

Methodologically, this method models neither situational nor action-formation mechanisms and accounts for transformational mechanisms via (manifest) aggregation (see Figure 2). Statistically, Croon and van Veldhoven (2007) have shown that this procedure only yields valid estimates if the L1 variance of the aggregated variables is zero. If the L1 variance is larger than zero, simple group mean aggregation yields biased estimates. In cross-national comparative survey research, this

1 In accordance with previous research, we use the terms 'conventional' or 'standard' multilevel analysis to describe hierarchical modeling techniques that are restricted to the analysis of level-one outcomes (Bennink, Croon, & Vermunt, 2013, 2015; Lüdtke et al., 2008; Preacher, Zyphur, & Zhang, 2010).

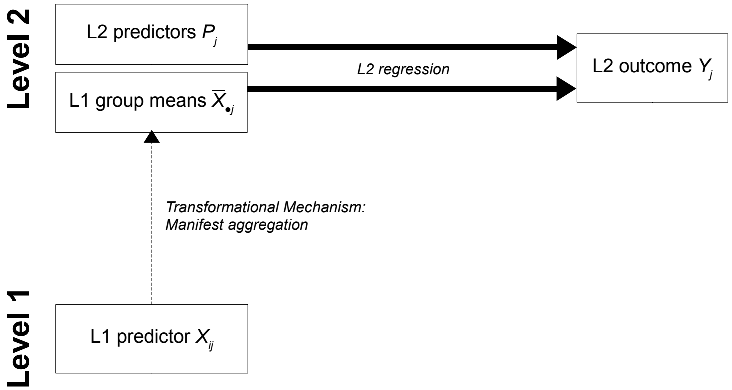

Figure 2 The simple group means approach

is often the case because individuals are sampled from a finite population and a specific constellation of individuals is selected to measure the L2 construct (Lüdtke et al., 2008). Since manifest aggregation does not control for these sampling errors, the observed group average (measured, for instance, in terms of group-specific arithmetic means) may be an unreliable measure of the unobserved true group average. In addition, the observed group average completely obscures the heterogeneity within groups. Therefore, if effects of observed group averages on L2 outcomes are of interest, estimates of both these effects and of other L2 predictors are likely to be biased when applying the simple group means approach (Bennink et al., 2013, 2015; Shin & Raudenbush, 2010).

The multilevel SEM approach

Multilevel SEM avoids these statistical problems by replacing manifest with latent aggregation (see Figure 3). Assume that we observe a manifest L1 variable X_{ij} for individuals i in countries j. X_{ij} is used to predict a manifest L2 outcome Y_j along with other L2 predictors P_j. Following the simple group means approach, X_{ij} is aggregated from L1 to L2 by computing group-specific arithmetic means $\overline{X_{\bullet j}}$, which are not corrected for sampling error. In a second step, $\overline{X_{\bullet j}}$ are used to predict Y_j controlled for P_j (adapted from Marsh et al., 2009):[2]

$$Y_j = \beta_0 + \beta_1 \overline{X_{\bullet j}} + \beta_2 P_j + u_{0j} \tag{1}$$

2 The notation by Marsh et al. (2009) implies group mean centering of all L1 predictors to account for a reference-group effect (in their example, this is the dependence of student academic self-concept on class-average achievement). Since our substantive application does not include a reference-group effect, we present the general notation without group mean centering. In addition, we use standard multilevel notation for the L2 residual variance.

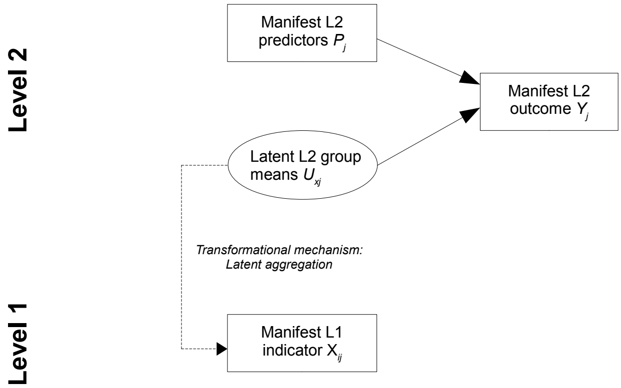

Figure 3 Latent aggregation in multilevel structural equation modeling

By contrast, multilevel SEM regards the actual group mean on L2 as an unobserved latent variable U_{xj} (which must not be confused with L2 residual error u_{oj}) that can only be estimated with error by the L1 indicators (Marsh et al., 2009). Following the conventions of SEM, the L2 latent means of the L1 observations are therefore depicted by ovals in Figure 3. While the simple group means approach treats the L2 group mean as a simple composite or index score of the L1 observations, multilevel SEM assumes the unobserved latent group means to *cause* the observed L1 values (Lüdtke et al., 2008).[3]

Multilevel SEM proceeds in two steps: First, an L2 latent variable U_{xj} is estimated. It is assumed to be the cause of X_{ij} at L1. In a second step, U_{xj} is used to predict the L2 outcome Y_j along with the other L2 predictors P_j:[4]

$$Y_j = \beta_0 + \beta_1 U_{xj} + \beta_2 P_j + u_{0j} \tag{2}$$

The aggregated L2 construct is a measure of the unobserved true group mean. Its reliability is a function of the relative share of the L2 variance weighted by the group-specific number of observations (Lüdtke et al., 2008):

$$\frac{\tau_x^2}{\tau_x^2 + \left(\sigma_x^2 / n_j\right)} \tag{3}$$

3 This points to the difference between formative and reflective models in measurement theory. Whereas formative latent variable models are already established in single-level measurement models (Diamantopoulos & Winklhofer, 2001), it remains unresolved whether formative latent aggregation is equally possible.

4 Additional controls for measurement error can be integrated easily (Marsh et al., 2009). For the sake of simplicity, our analysis of democratic persistence is limited to latent aggregation without controlling for measurement error.

As in conventional hierarchical modeling, σ^2_x denotes the L1 part and τ^2_x the L2 part of the variation of the respective indicator(s), whereas n_j refers to the group-specific number of observations.

By estimating a latent L2 variable U_{xj} as in (2), the variance of the L1 indicator is partitioned into an L1 and an L2 component. Unlike simple group mean aggregation, latent aggregation takes account of the heterogeneity within each group by partitioning the L1 variance σ^2_x from the L2 variance τ^2_x. In addition, by estimating latent group means at L2, which are assumed to cause the L1 observations in each group, the multilevel SEM approach acknowledges that the L1 scores do not perfectly map the construct at the L2 level, because of measurement error (Bennink et al., 2013, 2015; Preacher et al., 2010).

In sum, multilevel SEM replaces *manifest* with *latent* aggregation to aggregate individual-level predictors of macro-level outcomes. Like manifest aggregation, latent aggregation *per se* models only the transformational but not the situational and action formation mechanism. Statistically, however, latent aggregration is superior to manifest aggregation since it corrects for sampling error (see Table 1). As a result, its estimates are less biased, thereby permitting more valid inferences regarding the effect of multilevel predictors on contextual-level outcomes.

The two-step approach

The two-step approach also deals with the methodological and statistical issues that arise when studying multilevel models with contextual-level outcomes, albeit in a different manner. Figure 4 summarizes its basic idea.

The two-step approach builds on standard MLA. For an L1 outcome Y_{ij} and L1 units i nested in L2 contexts j, the standard model is given by:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \tag{4}$$

In equation (4), β_{0j} is the regression intercept of the outcome variable, β_{1j} is the regression slope of an L1 predictor, and e_{ij} is the residual error term. In contrast to non-nested regression analysis, both random intercepts β_{0j} and random slopes β_{1j} can be estimated for each L2 unit j by modeling them as a function of an additional L2 predictor Z_j with distinct intercepts (γ_{00} and γ_{10}) and regression slopes (γ_{01} and γ_{11}):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \tag{5}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \tag{6}$$

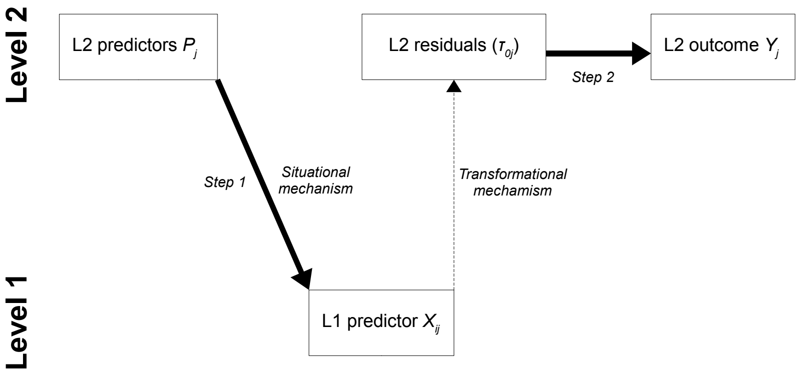

Figure 4 The two-step approach

Equations (5) and (6) introduce two additional residual error components: u_{0j} denotes the residual error of the outcome's L2 intercept β_{0j}, and u_{1j} denotes the residual error of the slope β_{1j} between L2 units.

Standard MLA only considers the case of an L1 outcome Y_{ij} that is predicted by L1 and L2 variables X_{ij} and Z_j, respectively. Griffin (1997) proposes an extension of the standard MLA approach to study an L2 outcome Y_j: Let X_{1ij} be the L1 explanatory variable of primary interest. In a first step, X_{1ij} is regressed on all other L1 and L2 predictors X_{2ij}, …, X_{nij} and Z_j:

$$X_{1ij} = \gamma_{00} + \gamma_{01} Z_j + \tau_{0j} + \beta_{1j} X_{2ij} + ... + \beta_{nj} X_{nij} + e_{ij} \tag{7}$$

In a second step, the L2 residuals u_{0j} of this model are used as a predictor variable in an L2 regression of the L2 outcome of interest:

$$Y_j = \beta_0 + \beta_1 u_{0j} + e_j \tag{8}$$

The effect of u_{0j} on the L2 outcome Y can be interpreted as the aggregated effect of the L1 variable X_1, net of both L1 and L2 covariates X_2, …, X_n and Z.

The two-step approach has both statistical and methodological advantages when studying multilevel models with contextual-level outcomes (see Table 1). Statistically, it provides a better estimate than the group mean aggregate: u_{0j} is a model-based estimate of the L2 variance that is already net of the L1 variance. In addition, u_{0j} can be adjusted for other covariates at L1 and L2. This may save degrees of freedom and circumvent collinearity issues when using u_{0j} as a predictor in a subsequent L2 regression. Compared to the group means approach and the multilevel SEM approach, the crucial methodological advantage of the two-step approach is its capacity to empirically model theoretical macro-micro-macro

Table 1 Comparison of methods for analyzing macro-micro-macro models

	Main methodological advantages & disadvantages	Main statistical advantages & disadvantages
group mean aggregation	Transformational mechanism (via manifest aggregation and macro regression)	Simple to perform, but only valid if variance of L1 variable = 0
ML SEM	Transformational mechanism (via latent aggregation and macro regression)	Takes sampling error into account: reduction of estimator bias
2-Step	1st step: situational & action-formation mechanism (via MLA) 2nd step: transformational mechanism (via residuals and macro regression)	Residual reflects the net effect of the individual-level independent variable

explanations in their entirety. The MLA of step 1 maps both the situational and action formation mechanism through the regression of an L1 outcome on L1 and L2 predictors. Storing the L2 residuals of this MLA then maps an underlying transformational mechanism in terms of an L1-L2 aggregation.

The relative statistical performance of each method can also be compared empirically. Based on previous research, we deduce two hypotheses. First, we expect that unless the L1 variance equals zero, simple group mean aggregation yields unreliable measures of the unobserved true group means. By contrast, multilevel SEM results in reliable estimates of true group means. Consequently, when group means based on simple aggregation are used as predictors of an L2 outcome, estimates of their regression coefficients may be biased (Bennink et al., 2013, 2015):

> H_1: Regression coefficients of L2 predictors that are simple group means deviate in terms of a) point estimates, b) standard errors, and c) resulting significance levels from regression coefficients of L2 predictors that have been aggregated through multilevel SEM.

Second, while the statistical performance of the two-step approach (Griffin, 1997) is less well researched, Lüdtke et al. (2008) compared multilevel SEM to another two-step approach proposed by Croon and van Veldhoven (2007). This approach adjusts the observed group means with weights from ANOVA formulas. This is quite similar to the decomposition of variance in an empty multilevel model. Lüdtke et al. (2008) observed that Croon and van Veldhoven's (2007) approach performed slightly less well than multilevel SEM. Consequently, we expect Griffin's two-step approach to yield estimates closer to multilevel SEM than to the simple group means approach:

> H_2: Regression coefficients of L2 predictors that have been aggregated by the two-step approach deviate less from multilevel SEM in terms of a) point estimates, b) standard errors, and c) resulting significance levels than regression coefficients of L2 predictors that are simple group means.

Substantive Application: A Multilevel Explanation of the Persistence of Democracy

Theoretical Background

To illustrate the methodological and statistical issues described in the previous section, we use the persistence of democracy as a substantive example. Explanations of democratic persistence pertain either to a macro-to-micro mechanism leading from the macro level to the level of individual citizens or to a micro-to-macro mechanism leading from individual citizens to the persistence of democracy at the macro level.

Przeworski (1991) introduces a classic model linking macro-level causes to individuals' micro-level incentives for subverting a democratic regime. Acknowledging that democratic competition produces winners and losers, he argues that "political forces comply with present defeats because they believe that the institutional framework that organizes the democratic competition will permit them to advance their interests in the future" (Przeworski, 1991, p. 19). Institutions are not only crucial for inspiring the belief that there will be future possibilities to advance one's interests. The given set of political and economic institutions also has distributional consequences affecting the capacities individuals have at their disposal to advance their interests (Przeworski, 1991). A model of democratic persistence therefore has to take into account that – under the same set of democratic rules – members of some societal groups might deem their chances of affecting future democratic outcomes to be lower than members of other societal groups. Correspondingly, classic studies have analyzed the decisive impact of economic development on both the process of successful democratization (Bollen, 1979; Bollen & Jackman, 1985; Lipset, 1959) as well as democratic persistence (Przeworski, Alvarez, Cheibub, & Limongi, 2000).

A second example for the macro-to-micro mechanism underlying the persistence of democracy is the idea that an ethnically divided society poses a particular challenge to democratic persistence (Horowitz, 1985; Rabushka & Shepsle, 1972; Reilly, 2001). In countries where several ethnic groups are politically mobilized, the question of who is to legitimately take part in the democratic game is continuously contested. Members of ethnic minorities often see little incentive to support ruling elites, who are – in virtue of the majority principle – likely to be members of the majority group. As a result, those out of power may choose to subvert democracy because they feel permanently excluded from democratic decisions likely to reflect only the interests of the majority.

A classic example of the micro-to-macro mechanism underlying the persistence of democracy is the political culture model. Almond and Verba (1963) seminally argued that the persistence of a political regime does not rest on its formal democratic institutions alone, but also on its political culture. Succeeding studies further specified the content of political culture and its effect on democratic persistence based on Easton's (1965, 1975) systems theory (Dalton, 2004; Fuchs, 2007; Norris, 1999). According to Easton, citizens' political support refers to their supportive values and attitudes toward the political community, the political regime, and political authorities (Easton, 1965). A critical amount of political support is necessary for any kind of political system to persist. Citizens' political support increases the functionality of political systems as it allows political authorities to convert demands into outputs and permits them to implement collectively binding decisions without having to resort to force (Easton, 1965).

Building on Easton (1965, 1975), Fuchs (2007) clarifies the implications of the different dimensions of political support for democratic political regimes. Support for the political authorities is crucial for their re- or de-election; support for the political system is essential for the persistence of a given type of democracy; support for democratic values is critical for the persistence of democracy in general (Fuchs, 2007). Thus, citizens' support for democratic values is the key factor when studying the effect of individual-level political orientations on the persistence of democracy at the macro level.

Fails and Pierce (2010) tested the systems approach of the political culture model empirically. Their analysis yielded no significant relationship between citizens' support for democratic values and their rejection of authoritarian values on the one hand and the probability of a decline of democracy on the other hand.

These mechanisms can be combined into a full multilevel explanation of democratic persistence (see Figure 5). From the macro to micro explanations, we take the insight that citizens' support for democratic values is likely to be affected by context-specific economic conditions and ethnic heterogeneity. From the micro to macro explanations, we take the insight that micro-level support for democratic values crucially accounts for the persistence of democracy at the macro level.

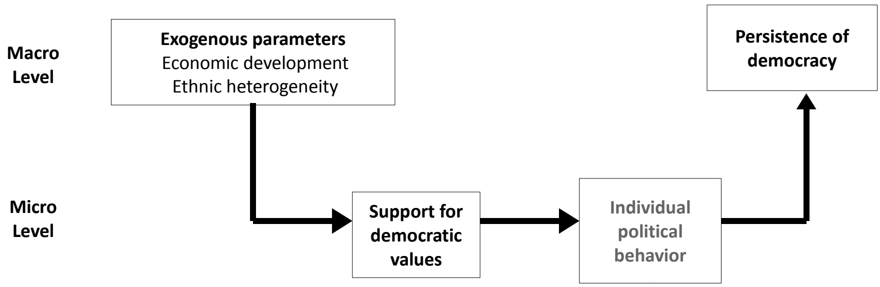

Figure 5 A two-level explanation of the persistence of democracy

Research Design

Period of analysis and data

Based on the data available, we analyzed the persistence of democracy from 1946 to 2014. We derived all L2 indicators from the *Quality of Government* standard time series data set (QoG) (Teorell et al., 2016), which includes data on a broad range of country-level indicators over time that we could easily merge with our L1 data.

To measure our L2 outcome variable – *democratic persistence* – we used the democracy index developed by the Polity IV project as included in the QoG (Marshall, Gurr, & Jaggers, 2015). Polity IV's democracy index – POLITY – reports countries' level of democracy on a scale ranging from -10 (fully autocratic) to +10 (fully democratic).[5] In line with the threshold provided on the Polity IV website (Marshall & Gurr, 2014), we considered countries as democracies if their POLITY score ≥ 6.[6]

As for our L2 predictors, we used the following indicators: *Economic development* was measured using countries' annual gross domestic product (GDP). We used the log of the OECD measure of GDP *per capita*. *Ethnic heterogeneity* was

5 POLITY is a composite score that quantifies the extent to which a country exhibits democratic and authoritarian characteristics. Polity IV coders assess countries' formal political institutions in terms of five component variables – the competitiveness of political participation (1), the openness of executive recruitment (2), the competitiveness of executive recruitment (3), the constraints on the executive (4), and the regulation of political participation (5) for each country on an annual basis. Countries are assigned weighted scores for each component. These are then added up to arrive at a democracy (DEMOC) and an autocracy score (AUTOC), both of which range from 0 to 10. The autocracy score is then subtracted from the democracy score to construct POLITY (Marshall et al., 2015).

6 We noted an inconsistency in the definition of the thresholds. In their codebook, Marshall et al. (2015) state that POLITY values ranging from +7 to +10 indicate a democratic regime.

measured using Fearon's (2003) ethno-linguistic fractionalization index (ELF), a measure of the probability that two randomly chosen individuals from a particular country are members of different ethnic groups. It ranges from 0 (perfect homogeneity) to 1 (very high fractionalization).[7]

Citizens' support for democratic values and all other L1 covariates were derived from the *World Values Survey* (WVS). The WVS is a cross-national survey based on representative national samples investigating worldwide socio-cultural and political change. For our analyses, we used the wave 6 aggregated longitudinal file, which includes more than 340,000 observations sampled in 101 countries across all available waves from 1981 to 2014. In line with previous research, *support for democratic values* was operationalized in terms of respondents' reply to the following question: "I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good, fairly good, fairly bad or very bad way of governing this country?". For reasons of data availability, we used respondents' rejection of an authoritarian system rather than their support for a democratic system. The answer category reads: "Having a strong leader who does not have to bother with parliament and elections" (1 = 'very good'; 2 = 'fairly good'; 3 = 'bad'; 4 = 'very bad'). For our analyses, we dichotomized this variable (0 = 'good / very good' vs. 1 = 'bad / very bad'). In accordance with previous research (Schneider, 2009), we controlled for individuals' age (six categories ranging from 1 = '15-24 years' to 6 = '65 and more years'), subjective assessment of social class (five categories ranging from 1 = 'lower class' to 5 = 'upper class'), and education (eight categories ranging from 1 'inadequately completed elementary education' to 8 'university with degree/higher education').[8]

Methods of analysis

Studying the effect of L1 and L2 predictors on an L2 outcome such as the persistence of democracy poses two methodological challenges. First, choosing a method to address the L1-L2 aggregation problem; second, analyzing persistence of democracy, which is a duration variable.

We compared three different strategies for solving the L1-L2 aggregation problem. First, we aggregated support for democratic values and all other L1 covariates by computing the arithmetic means for each country year (model 1). Second, we corrected for sampling error by estimating a latent aggregation of all L1 variables on L2 using multilevel SEM (model 2).[9] Third, we applied the two-step procedure

7 The formula is: $1 - \sum_{i=1}^{n} s_i^2$ where s_i is the share of group i ($i = 1, \ldots, n$).

8 See Table A1 (appendix) for a summary of all variables.

9 The latent aggregation was performed in Mplus, Version 7 (Muthén & Muthén, 2012).

proposed by Griffin (1997) by regressing support for democratic values on all other L1 and L2 predictors and then using the L2 residuals of this multilevel model as a new predictor variable.

We estimated not one, but several multilevel levels that were built up stepwise: The first empty model separated the L2 residuals of support for democratic values from the L1 residuals (model A1). We then added the macro level predictors GDP and ELF (models A2-A4). Finally, we added all L1 controls (model A5).[10] Researchers typically use stepwise model building (which we also carried out in the L2-only regressions below) to make causal claims about mediator variables partialing out significant effects of previous regressors. Apart from comparing point estimates and confidence intervals between aggregation methods for the final model, we also considered it instructive to analyze a series of stepwise models in order to assess whether different aggregation methods lead to different claims about causal mediation.

In addition, we chose an adequate model for predicting democratic persistence, a duration variable. The time span of interest is the persistence of a given democracy until its breakdown. Whereas some democracies may have persisted before entering the observation window (left censoring), others may have continued to persist after the observation ended (right censoring). Within the time period of analysis, the same country may have experienced multiple democratic sequences, followed by breakdowns. In order to address these issues, we used event history modeling. We considered democratic breakdown to occur if the score of democratic regimes (nested within countries) fell below the threshold of POLITY = 6. The duration until this event was measured by the total number of years a democratic system persisted from 1946 onwards. Multiple breakdowns within the same country were coded as distinct events. To keep the models parsimonious, we used a simple exponential event history model, which assumes constant transition rates across years.

In formal terms, our event history model is defined as follows: Let h denote the hazard rate of democracies' estimated risk of falling below POLITY = 6 and t the time of democracies' survival. The basic exponential survival model can then be described as:

$$h(t) = \lambda; t > 0, \lambda > 0 \tag{9}$$

λ is a positive constant constraining transition rate (in terms of democratic breakdowns) that is equal across years. Our aim was to predict the expected survival time E(t) with an aggregate measure of citizens' support for democratic values (*DVAL*),

10 See Table A2 (appendix).

countries' *GDP* and *ELF,* as well as aggregate measures of citizens' age (*AGE*), subjective social class (*SCLASS*), and education (*EDUC*).

When applying simple aggregation, democracies' expected time of survival was estimated by:

$$E\left(t_j\right) = \exp\begin{pmatrix} \beta_0 + \beta_1 \overline{DVAL_{\bullet j}} + \beta_2 GDP_j + \beta_3 ELF_j + \beta_4 \overline{AGE_{\bullet j}} \\ + \beta_5 \overline{SCLASS_{\bullet j}} + \beta_6 \overline{EDUC_{\bullet j}} \end{pmatrix} \tag{10}$$

where $\overline{X_{\bullet j}}$ from equation (1) was replaced by the aforementioned predictor variables. When using latent aggregation, we estimated:

$$E\left(t_j\right) = \exp\begin{pmatrix} \beta_0 + \beta_1 U\left(DVAL\right)_j + \beta_2 GDP_j + \beta_3 ELF_j + \beta_4 U\left(AGE\right)_j \\ + \beta_5 U\left(SCLASS\right)_j + \beta_6 U\left(EDUC\right)_j \end{pmatrix} \tag{11}$$

Here, *U* refers to the unobserved latent L2 group mean which is assumed to cause the observed L1 values of each variable.

Finally, when employing the two-step approach, the estimates were derived as follows:

$$E\left(t_j\right) = \exp\left(\beta_0 + \beta_1 u_{0jm}\right) \tag{12}$$

In equation (12), u_{0jm} denotes the L2 residuals from a hierarchical regression of citizens' support for democratic values on both the L2 predictors and the L1 covariates. The subscript *m* indicates that the hierarchical models were built up in a stepwise manner, which is why we estimated several terms for u_0.

These formal specifications require a methodological addendum: While we *estimated* three *L2* event history analyses after having applied each of the three aggregation methods, our *theoretical* explanation emphasizes the importance of citizens' support for democratic values on *L1*. Hence, though the event history models applied L2-only regressions, in line with the paradigm of structural individualism, we assume that the theoretical mechanisms operate via citizens' preferences and beliefs on the micro level. In line with the aim of our article, we sought to determine how the three different aggregation methods map these L1 processes when predicting an L2 outcome.

In order to increase our statistical power, we used both inter- and extrapolation techniques for our independent variables. We interpolated missing values between observation points, using the *-ipolate-* command in Stata. In addition, we extrapolated missing values between the last valid observation and 2015, using a 'non-linear trend' scenario. We first estimated a polynomial regression of the interpolated values of each predictor on years of observations using the *-lpoly-* command in

Stata. We then used out-of-sample predicted values to replace missing observation for subsequent years over countries.[11]

Results

Prior to computing the comprehensive multivariate models, we compared the survival functions of democracies with high vs. low average support for democratic values. We dichotomized the support variable and compared countries with one standard deviation above vs. below the grand mean of the aggregated variable. We then compared the survival functions of these two groups of countries using group mean aggregation, the two-step model, and latent aggregation. Independent of the method of aggregation, in the long run, the estimated survival rate for democracies scoring one standard deviation above the grand mean of support for democratic values was higher than for their lower-scoring counterparts (see Figure A3, appendix). Apart from a lower estimate of the survival rate of countries whose citizens had less support for democratic values in the two-step model, the differences between the aggregation methods appeared to be negligible.

Figure 6 presents the results of the analyses using the simple group means approach (model 1), multilevel SEM (model 2), and the two-step approach (model 3). It shows both point estimates and confidence intervals for the L1 and L2 predictors. Our survival models were built up stepwise: In model 1a and 2a, the survival rate of democracies was first predicted by support for democratic values only; in model 3a, it was predicted by the L2 residuals from the multilevel null model, which separated the variance of the L1 support variable without having included any other L1 or L2 predictor. In models 1b and models 2b, we simultaneously added GDP and ELF. Correspondingly, in model 3b we included the residuals corrected for these L2 predictors. Finally, in model 1c and 2c, we added the L1 covariates; in model 3c we included the residuals corrected for the L1 covariates. Because of the low number of events, we displayed confidence intervals both on the 10% ($|t| > 1.64$; see ticks of confidence bands) and the 5% significance level ($|t| > 1.96$; see ends of confidence bands).

When applying the *simple group means approach*, support for democratic values did not turn out to be a significant predictor of democratic survival. Point estimates varied between -3.734 in model 1a and -3.367 in model 1c, but neither

11 The overlap of valid observations for both democratic persistence and support for democratic values before and after interpolation is displayed in Figure A1 (appendix). The basic survivor function of democratic persistence for our reduced sample of analysis is sufficiently similar to the survivor function of the total country sample (see Figure A2, appendix). As a sensitivity check, we also extrapolated our interpolated values by repeating the last valid observation of each predictor for subsequent years with missing values. Results based on this extrapolation technique are very similar to the results reported in the results section (see Figure A4, appendix).

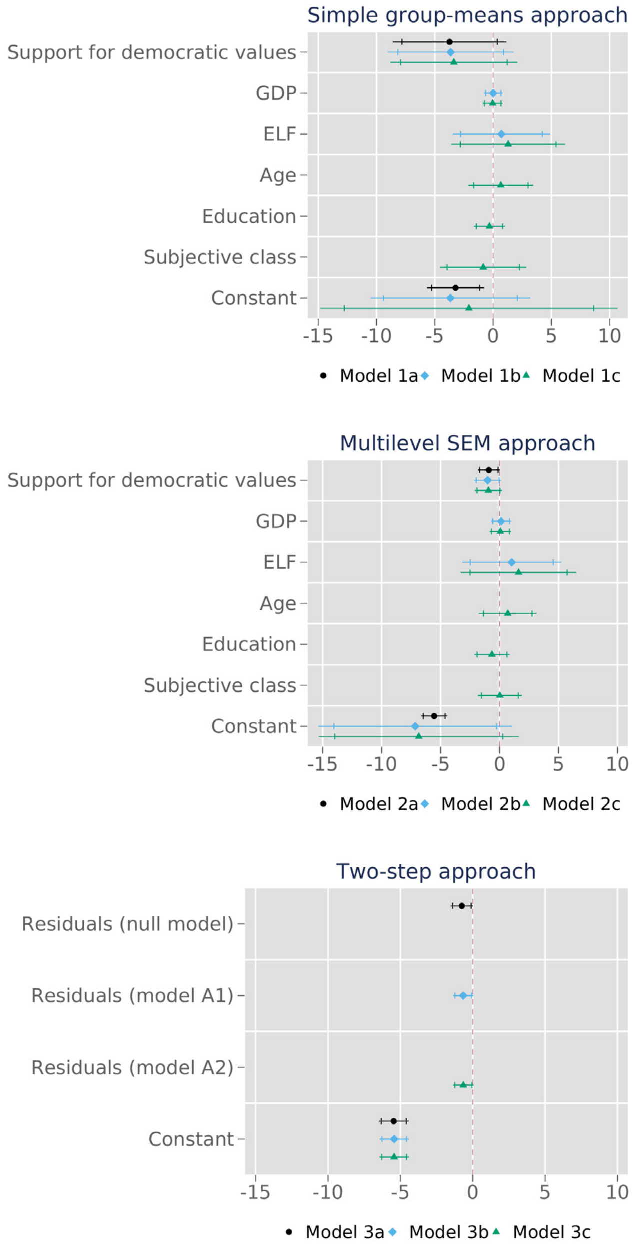

Figure 6 Point estimates and confidence intervals of countries' democratic survival across aggregation methods. *N*= 917 observations, *N*= 122 subjects, *N*= 5 failures in all models

estimate was larger than 1.65 times its standard error (also see Table A3, appendix). The latter also applies to all other L2 predictors and to the L1 covariates. We observed significant intercept variation in model 1a, which only included support for democratic values as a predictor variable, but not in models 1b and 1c, which controlled for the remaining variables. Values of AIC and BIC as indicators of model fit show that not much was gained by adding predictors of democratic survival apart from citizens' support for democratic values (see Table A3, appendix).

When using the *latent aggregation approach*, the estimated confidence intervals of support for democratic values became more precise and we observed two effects of support for democratic values on democratic survival that were greater than 1.65 times their standard error (models 2a and 2b). Once the aggregated L1 covariates were controlled for, our predictor was no longer significantly associated with the outcome. Point estimates were remarkably lower after latent aggregation, ranging from -.911 in model 2a to -1.009 in model 2b (see Table A4, appendix). Having controlled for L2 structural conditions (in terms of GDP and ELF), the effect of support for democratic values became more negative from model 2a to model 2b – which points to a suppressor effect. Yet, similar to the simple group means analysis, none of the remaining variables turned out to be significant predictors of democratic survival. Model fit indices again supported the most parsimonious model 2a and intercept variation was significant in the first two submodels only.

When applying the *two-step approach*, point estimates of support for democratic values on democratic stability were predicted with similar precision as in latent aggregation when looking at the confidence intervals. Yet, in the two-step model, we observed three significant effects at the 10% level. The L2 (u_{oj}) residuals of support for democratic values predicted democratic survival independent of whether they were adjusted for other L1 or L2 variables. Effect sizes ranged from -.754 in model 3a to -.651 in model 3c (see Table A5, appendix). In contrast to simple group mean and latent aggregation, the intercept remained significant in all three sub-models. Though model fit indices supported the most parsimonious model 3a, the differences between model fit indices across models were less striking than in the event history regressions following manifest and latent aggregation.

Our results can be summarized as follows: In each estimation, support for democratic values was negatively associated with the event of democratic breakdown, as expected by theory. This replicated our bivariate analysis where democracies with higher support for democratic values showed a longer estimated survival rate on average. Apart from this similarity, there are notable differences between the aggregation methods: While support for democratic values was not significantly associated with democratic stability after manifest aggregation, significant effects could be observed after both latent aggregation and the two-step approach. Applying more advanced aggregation methods led to smaller point estimates and standard errors compared to the simple group means approach. All this is in line with the two hypotheses postulating notable differences between simple group means aggregation and latent aggregation, and closer similarity between the two-step approach and latent aggregation than between the two-step approach and manifest aggregation.

Yet, compared to latent aggregation, which has already been observed to yield unbiased point estimates in simulation models (Bennink et al., 2013, 2015; Lüdtke et al., 2008), researchers who apply the two-step approach may run the risk of committing type one errors: In the most comprehensive model of the two-step approach (model 3c) and unlike in the corresponding regressions following latent aggregation (model 2c), the effect of support for democratic values was significant at the 10% level.[12]

Conclusion

In this paper, we addressed a methodological challenge well known to comparative survey researchers: how to study the effect of level two (L2) and level one (L1) predictors of a level two (L2) outcome so as to yield both reliable and valid results. Researchers have criticized simple aggregation for methodological and statistical reasons. Building on these insights and using the persistence of democracy as a substantive example, we compared the simple group means approach with two more advanced analytical strategies: the multilevel SEM approach, which estimates a latent L2 variable assumed to cause its L1 indicators, and a two-step approach, which relies on the L2 residuals of a multilevel model estimated prior to the analysis of interest (Griffin, 1997).

Our study corroborates previous critiques of the simple group-means approach. In both bivariate comparisons of countries' survival curves and more comprehensive multivariate event history analyses, we observed that support for democratic values was negatively associated with democratic breakdown. Unlike in the bivariate models, however, the multivariate models revealed that the associated significance levels of the estimates of support for democratic values differed remarkably depending on the aggregation method. Whereas support for democratic values was not significant in the regressions following simple group mean aggregation, confidence intervals suggested point estimates of higher precision when using either the multilevel SEM or the two-step approach, and the latter two approaches showed several significant effects at the 10% level.

These empirical results show that researchers can improve the validity of their inferences by choosing more advanced analytical strategies. First, the results match previous findings from simulation analyses (Lüdtke et al., 2008), which show that

12 The event-history models underlying Figure 6 are listed in Tables A3 to A5 (appendix).

the simplest form of aggregation – manifest group means – is prone to beta or type-two errors in terms of false negative findings. Second, our results challenge Fails and Pierce's (2010) finding (based on simple aggregation) that support for democratic values has no effect on democracies' probability of decline. Our results suggest that comparative survey researchers interested in the effect of one or more L1 predictors on an L2 outcome may overestimate the standard errors of their regression coefficients when using manifest group mean aggregation.

The two more advanced analytical strategies have distinct methodological and statistical advantages. From a statistical perspective, the two-step approach performs somewhat poorer than the multilevel SEM approach: Given that simulation revealed regression coefficients after latent aggregation to be unbiased (Bennink et al., 2013, 2015; Lüdtke et al., 2008), researchers who apply the two-step approach may run the risk of committing type-one errors in terms of false positive findings. An evident methodological advantage of the two-step approach is, however, that it is particularly suited to simultaneously model situational, action formation, and transformational mechanisms in their entirety.

We conclude with several suggestions for future research. As of yet, no simulation analyses (similar to the ones comparing the simple group mean and the multilevel SEM approach) have been carried out for the two-step approach. It is therefore not possible to determine whether the estimated confidence intervals of the two-step approach are more or less reliable than the results of the latent aggregation approach. Hence, our first suggestion for future research is to perform a simulation analyses for all three aggregation methods. Controlling the data-generating mechanism would permit valid conclusions about the actual precision of each aggregation method compared to the 'real' effect size at L2.

Second, the latent aggregation model can be extended towards a *doubly-latent* model with controls for measurement error. Thus, our second suggestion for future research is to use multiple indicators of political support to arrive at a doubly-latent model of political support at L2. Depending on the results of the aforementioned simulation study, latent variable models and the two-step approach could eventually also be combined in order to estimate both situational and transformational mechanisms without falling prey to either measurement or sampling error. Moreover, if individuals' actual decisions such as turning out to vote or participating in demonstrations or public protests are considered, a combined framework of structural equation modeling and the two-step approach would allow researchers to map action-formation mechanisms as well.[13] Third, while we used a simple exponential event-history model to simplify the analysis, future research might make use of

---

13 Structural equation modeling can map action formation mechanisms in simple L1 regressions as well. In addition, for group-mean centered L1 variables, multilevel SEM can estimate situational mechanisms by computing the difference between L2 and L1 regression coefficients (Marsh et al., 2009).

more flexible links for the survival function such as piecewise constant or frailty models.

In sum, we encourage comparative survey researchers to surpass the simple group means aggregation approach in favor of more advanced methods of analyzing contextual-level outcomes. We have shown that this helps researchers to circumvent beta or type-two errors in terms of false negative findings when using one or more L1 indicator to predict an L2 outcome. In addition, unlike the simple group means approach, these more advanced methods can be extended further, thereby facilitating the test of more theoretically valid models.

References

Almond, G. A., & Verba, S. (1963). *The civic culture: Political attitudes and democracy in five nations*. Princeton, NJ: Princeton University Press.

Becker, D., Beckers, T., Franzmann, S. T., & Hagenah, J. (2016). Contextualizing cognitive consonance by a social mechanisms explanation: Moderators of selective exposure in media usage. *Analyse & Kritik*, *38*(1), 149-178. http://doi.org/10.1515/auk-2016-0108

Bennink, M., Croon, M. A., & Vermunt, J. K. (2013). Micro-macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. *Sociological Methods & Research*, *42*(4), 431-457. https://doi.org/10.1177/0049124113500479

Bennink, M., Croon, M. A., & Vermunt, J. K. (2015). Stepwise latent class models for explaining group-level outcomes using discrete individual-level predictors. *Multivariate Behavioral Research*, *50*(6), 662-675. http://dx.doi.org/10.1080/00273171.2015.1074879

Bollen, K. A. (1979). Political democracy and the timing of development. *American Sociological Review*, *44*(4), 572-587. Retrieved from http://www.jstor.org/stable/2094588

Bollen, K. A., & Jackman, R. W. (1985). Economic and noneconomic determinants of political democracy in the 1960s. In R. G. Braungart & M. M. Braungart (Eds.), *Research in political sociology* (pp. 27-48). Greenwich, CT: JAI Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*(1), 45-57. http://doi.org/10.1037/1082-989X.12.1.45

Dalton, R. J. (2004). *Democratic challenges, democratic choices: The erosion of political support in advanced industrial democracies*. Oxford, UK: Oxford University Press.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269-277.

Easton, D. (1965). *A systems analysis of political life*. New York, NY: Wiley.

Easton, D. (1975). A re-assessment of the concept of political support. *British Journal of Political Science*, *5*(4), 435-457.

Fails, M. D., & Pierce, H. N. (2010). Changing mass attitudes and democratic deepening. *Political Research Quarterly*, *63*(1), 174-187. http://doi.org/10.1177/1065912908327603

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, *8*(2), 195-222. http://doi.org/10.1023/A:1024419522867

Fuchs, D. (2007). Political culture paradigm. In R. J. Dalton & H.-D. Klingemann (Eds.), *The Oxford handbook of political behavior* (pp. 161–184). Oxford, UK: Oxford University Press.

Griffin, M. A. (1997). Interaction between individuals and situations: Using HLM procedures to estimate reciprocal relationships. *Journal of Management*, *23*(6), 759-773. http://doi.org/10.1016/S0149-2063(97)90028-3

Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge, UK: Cambridge University Press.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, *36*, 49-67. https://doi.org/10.1146/annurev.soc.012809.102632.

Horowitz, D. L. (1985). *Ethnic groups in conflict*. Berkeley, CA: University of California Press.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Lim, F., Bond, M. H., & Bond, M. K. (2005). Linking societal and psychological factors to homicide rates across nations. *Journal of Cross-Cultural Psychology*, *36*(5), 515-536. http://doi.org/10.1177/0022022105278540

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *The American Political Science Review*, *53*(1), 69-105. http://doi.org/10.2307/1951731

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229. http://doi.org/10.1037/a0012869

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764-802. http://doi.org/10.1080/00273170903333665

Marshall, M. G., & Gurr, T. R. (2014). *Polity IV Project*. Retrieved from http://www.systemicpeace.org/polity/polity4.htm

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2015). *Polity IV Project: Regime characteristics and transitions, 1800-2015. Dataset users' manual*. Vienna, Austria: Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/inscr/p4manualv2015.pdf

Muller, E. N., & Seligson, M. A. (1994). Civic culture and democracy: The question of causal relationships. *American Political Science Review*, *88*(3), 635-652. http://doi.org/10.2307/2944800

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Norris, P. (Ed.). (1999). *Critical citizens: Global support for democratic government*. Oxford, UK: Oxford University Press.

OECD. (2014). *Social expenditure: Reference series (Edition 2016)*. [Data file]. Retrieved from http://doi.org/10.1787/socx-data-en

Opp, K.-D. (2011). Modeling micro-macro relationships: Problems and solutions. *Journal of Mathematical Sociology*, *35*(1-3), 209-234. http://doi.org/10.1080/0022250X.2010.532257

Preacher, K. J., Zyphur M. J., & Zhang, Z. 2010. A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods 15*(3), 209-233. http://psycnet.apa.org/doi/10.1037/a0020141

Przeworski, A. (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America*. Cambridge, UK: Cambridge University Press.

Przeworski, A., Alvarez, M. E., Cheibub, J. A., & Limongi, F. (2000). *Democracy and development: Political institutions and well-being in the world, 1950-1990*. Cambridge, UK: Cambridge University Press.

Rabushka, A., & Shepsle, K. A. (1972). *Politics in plural societies: A theory of democratic instability*. Columbus, OH: Merrill.

Raub, W., Buskens, V., & van Assen, M. A. L. M. (2011). Micro-macro links and microfoundations in sociology. *The Journal of Mathematical Sociology*, *35*(1-3), 1-25. http://doi.org/10.1080/0022250X.2010.532263

Reilly, B. (2001). *Democracy in divided societies: Electoral engineering for conflict management*. Cambridge, UK: Cambridge University Press.

Schneider, C. Q. (2009). *The consolidation of democracy: Comparing Europe and Latin America*. New York, NY: Routledge.

Shin, Y. & Raudenbush, S. W. 2010. A Latent Cluster-Mean Approach to the Contextual Effects Model With Missing Data. *Journal of Educational and Behavioral Statistics, 35*(1), 26-53. https://doi.org/10.3102/1076998609345252

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A., & Svensson, R. (2016). *The quality of government standard dataset (Version Jan16)*. [Data file]. University of Gothenburg: The Quality of Government Institute. http://doi.org10.18157/QoGStdJan16

Tranow, U., Beckers, T., & Becker, D. (2016). Explaining and understanding by answering "why" and "how" questions: A programmatic introduction to the special issue "social mechanisms." *Analyse & Kritik*, *38*(1), 1-30. http://doi.org/10.1515/auk-2016-0102

Udehn, L. (2002). The changing face of methodological individualism. *Annual Review of Sociology*, *28*, 479-507. Retrieved from http://www.jstor.org/stable/3069250

World Values Survey Association. (2015). *World Values Survey 1981-2014 Longitudinal Aggregate (v. 20150418)*. [Data file]. Retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp

Appendix

Table A1 Distribution of all indicators

		count	mean	sd	min	max
LEVEL 1	Support for democratic values	269869	2.75	1.03	1	4
	Support for democratic values (dichotomized)	269869	0.59	0.49	0	1
	Age recoded	337018	3.1	1.57	1	6
	Highest educational level attained	296142	4.72	2.23	1	8
	Subjective social class	284337	2.68	0.99	1	5
LEVEL 2	GDP	7998	7.62	1.64	3.51	12.11
	ELF	8573	0.47	0.27	0.00	1.00
	Support for democratic values	1007	0.60	0.18	0.01	0.97
	Age	1190	3.19	0.46	1.91	4.30
	Education	1076	4.74	0.80	2.53	6.79
	Subjective class	1022	2.69	0.28	1.70	3.69
	Residuals (null model)	921	-0.02	0.86	-4.84	3.06
	Residuals (model A1)	921	0.00	0.93	-5.34	2.89
	Residuals (model A2)	921	-0.01	0.95	-5.30	2.94
	Support for democratic values	1007	-0.02	0.82	-3.89	2.22
	Age	1058	0.07	0.53	-1.47	1.32
	Education	1034	0.09	0.65	-1.67	1.71
	Subjective class	1013	0.04	0.51	-1.67	1.35

Table A2 Multilevel logistic regression of support for democratic values (dichotomized) on level-two predictors and level-one covariates

	Null model		Model 1a		Model 1b	
	b	*se*	*b*	*se*	*b*	*se*
Intercept	1.812**	(0.585)	2.042***	(0.580)	0.457***	(0.071)
log(GDP)			-0.174**	(0.061)	-0.166**	(0.061)
ELF			-0.392	(0.345)	-0.427	(0.341)
Age: 15-24 years	REFERENCE CATEGORY					
25-34 years					0.015	(0.015)
35-44 years					0.067***	(0.015)
45-54 years					0.103***	(0.017)
55-64					0.092***	(0.019)
65 and more years					-0.039	(0.020)
Education: Inadequately completed elementary	REFERENCE CATEGORY					
Completed elementary					0.042	(0.022)
Incomplete secondary: tech./voc.					0.051*	(0.025)
Completed secondary: tech./voc.					0.178***	(0.022)
Incomplete secondary: univ. prep.					0.171***	(0.024)
Complete secondary: univ. prep.					0.274***	(0.022)
Some university without degree					0.428***	(0.026)
University with degree					0.581***	(0.023)
Subjective class: lower	REFERENCE CATEGORY					
working					0.016	(0.017)
lower middle					0.042*	(0.017)
upper middle					-0.034	(0.019)
upper					-0.275***	(0.038)
τ_{0j}	0.025	(0.063)	0.012	(0.063)	-0.045	(0.054)
N	219740		219740		219740	
AIC	261954		263445		263440	

Notes. Random intercept model (QR decomposition) across country-years (level 2). Significance levels: * < .05; ** < .01; *** < .001 (two-sided). Standard errors in parentheses.

Table A3 Exponential event-history regression of democratic breakdown on aggregated support for democratic values, L2 predictors, and aggregated L1 controls (simple group-means approach)

	Model 2a *b/se*	Model 2b *b/se*	Model 2c *b/se*
Intercept	-3.220*	-3.662	-2.073
	(1.252)	(3.492)	(6.503)
Support for democratic values	-3.734	-3.642	-3.367
	(2.485)	(2.754)	(2.783)
log(GDP)		0.01	-0.038
		(0.399)	(0.432)
ELF		0.715	1.294
		(2.131)	(2.495)
Age			0.662
			(1.419)
Education			-0.315
			(0.685)
Subjective class			-0.846
			(1.887)
AIC	43.318	47.201	52.375
BIC	52.96	66.486	86.123
N (failures)	5	5	5
N (subjects)	122	122	122
N (observations)	917	917	917

Notes. Significance levels: + < .10: * < .05; ** < .01; *** < .001 (two-sided). Standard errors in parentheses.

Table A4 Exponential event-history regression of democratic breakdown on aggregated support for democratic values, L2 predictors, and aggregated L1 controls (multilevel SEM approach)

	Model 3a	Model 3b	Model 3c
	b/se	*b/se*	*b/se*
Intercept	-5.547***	-7.151+	-6.851
	(0.563)	(4.195)	(4.332)
Support for democratic values	-0.911+	-1.009+	-0.945
	(0.474)	(0.592)	(0.591)
GDP		0.132	0.064
		(0.428)	(0.461)
ELF		1.029	1.611
		(2.141)	(2.502)
Age			0.696
			(1.249)
Education			-0.644
			(0.769)
Subjective class			0.024
			(0.949)
AIC	42.444	46.179	51.203
BIC	52.086	65.463	84.951
N (failures)	5	5	5
N (subjects)	122	122	122
N (observations)	917	917	917

Notes. Significance levels: + < .10: * < .05; ** < .01; *** < .001 (two-sided). Standard errors in parentheses.

Table A5 Exponential event-history regression of democratic breakdown on residualised support for democratic values (two-step approach)

	Model 4a *b/se*	Model 4b *b/se*	Model 4c *b/se*
Intercept	-5.460*** (0.525)	-5.427*** (0.517)	-5.433*** (0.520)
Residuals (Null model)	-0.754+ (0.389)		
Residuals (model 1a)		-0.658+ (0.357)	
Residuals (model 1b)			-0.651+ (0.361)
AIC	42.813	43.047	43.089
BIC	52.455	52.689	52.731
N (failures)	5	5	5
N (subjects)	122	122	122
N (observations)	917	917	917

Notes. Significance levels: + < .10: * < .05; ** < .01; *** < .001 (two-sided). Standard errors in parentheses.

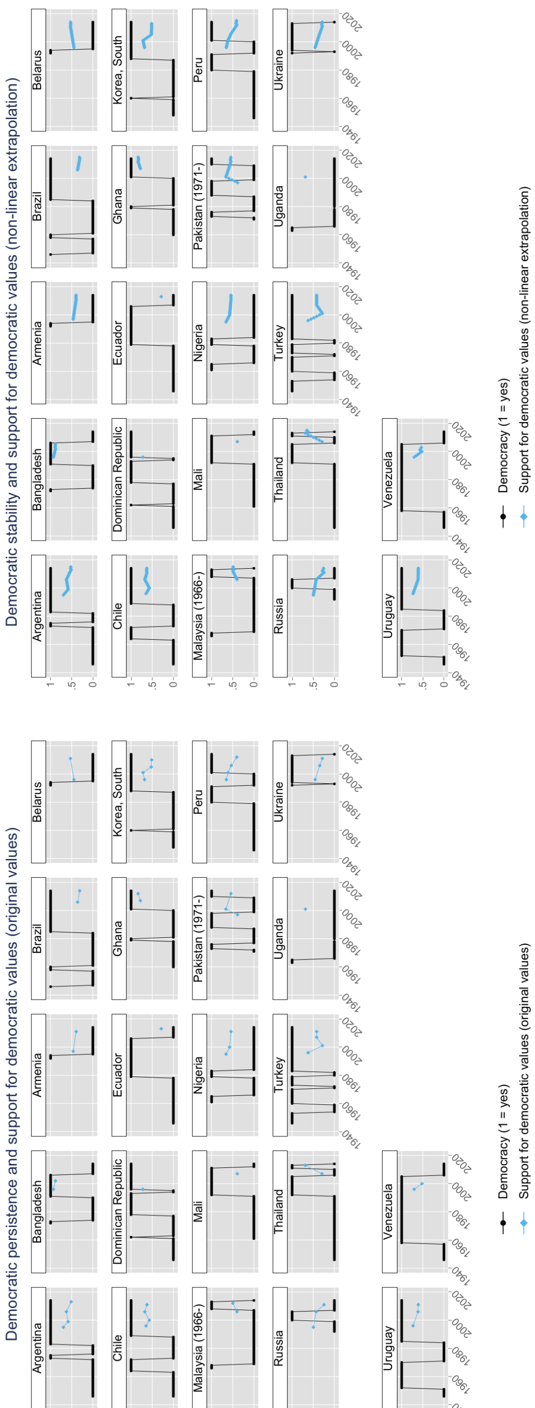

Figure A1 Distribution of democratic persistence and support for democratic values across country years

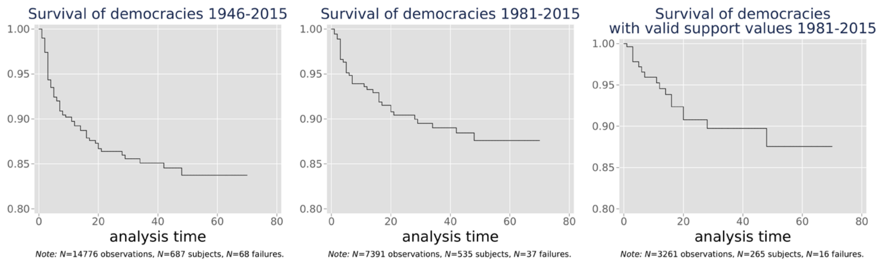

Figure A2 A comparison of democracies' estimated survival rates across different samples of analysis

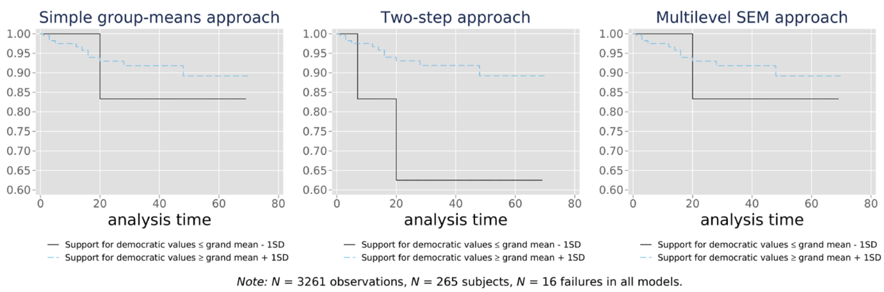

Figure A3 Survival of democracies by support for democratic values across aggregation methods

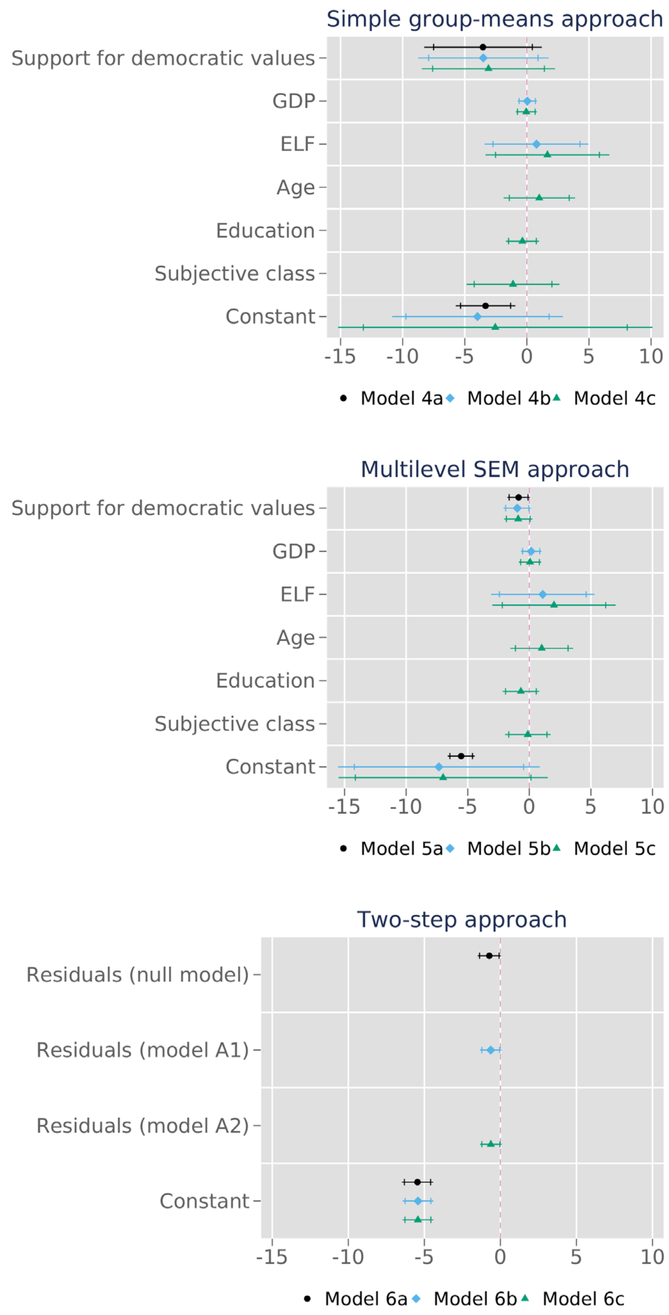

Figure A4 Point estimates and confidence intervals of countries' democratic survival across aggregation methods (constant interpolation)